Counting Reads and Differential Expression

Short Read Workshop
Day 7

Recap of Day 7 Videos

- Counting reads (for RNA-seq)
  - Steps to counting reads and normalizing counts
- Differential expression overview
  - Introduction to differential expression analysis
- Differential expression analysis with DESeq2
  - DESeq2 Negative Binomial theoretical model
- Multifactor designs with DESeq2
  - Expanded designs in DESeq2

Goal of the Day

Find genes that are different between samples

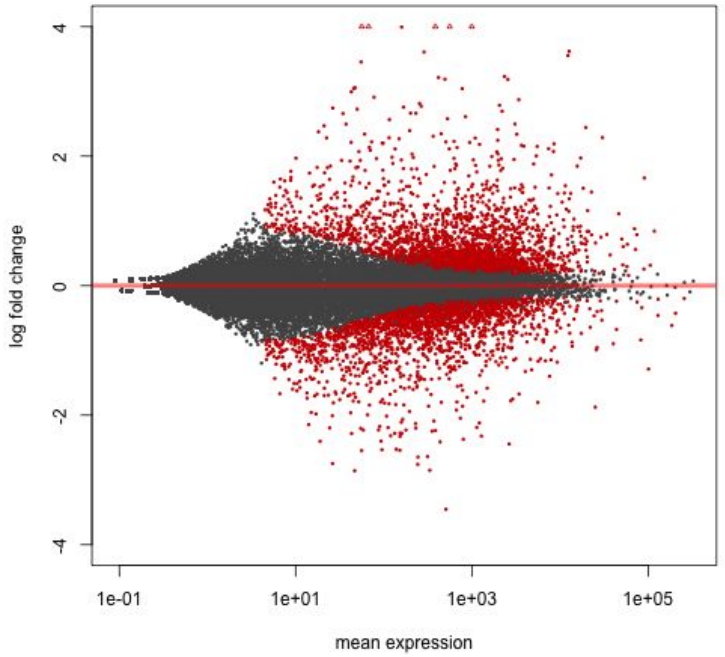
Map reads to reference genome

↓

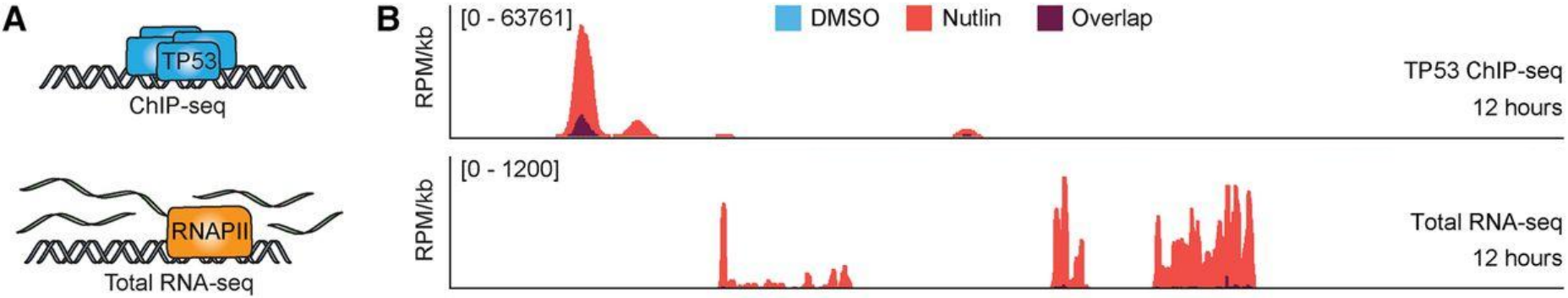
Count reads

↓

Perform differential gene expression analysis



Project B: Identification of the p53 transcriptional program using RNA-seq and ChIP-seq

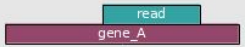
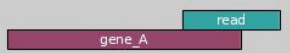
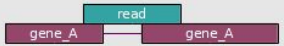
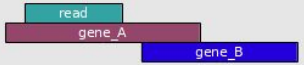
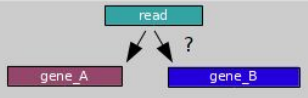


Andrysik et al., 2017, doi: 10.1101/gr.220533.117

In HCT116, SJSA, and MCF7 cell types.

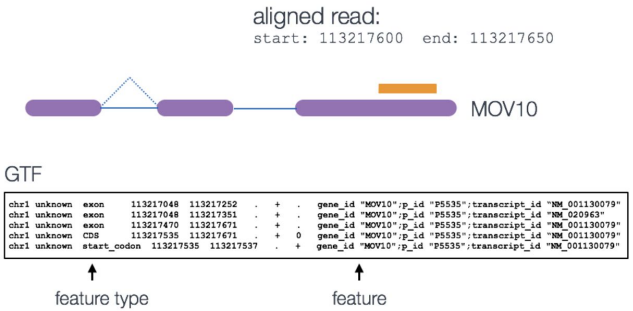
Question: Which genes are driven by p53 activation?

`featureCounts` counts reads over features in R

	union	intersection_strict	intersection_nonempty
	gene_A	gene_A	gene_A
	gene_A	no_feature	gene_A
	gene_A	no_feature	gene_A
	gene_A	gene_A	gene_A
	gene_A	gene_A	gene_A
	ambiguous (both genes with --nonunique all)	gene_A	gene_A
	ambiguous (both genes with --nonunique all)		
	alignment_not_unique (both genes with --nonunique all)		

There are several options in `featureCounts`

```
fc <- featureCounts(files=bam_file_list,
                    annot.ext=gtf,
                    isGTFAnnotationFile=TRUE,
                    GTF.featureType="exon",
                    GTF.attrType="gene_id",
                    useMetaFeatures=TRUE,
                    allowMultiOverlap=TRUE,
                    largestOverlap=TRUE,
                    countMultiMappingReads=TRUE,
                    isPairedEnd=TRUE,
                    strandSpecific=1,
                    nthreads=N)
```



https://hbctraining.github.io/Intro-to-rnaseq-hpc-O2/lessons/05_counting_reads.html

Counting reads with `featureCounts`

- Follow featureCounts worksheet:
  - Open `R` and install `Rsubread on AWS`
  - Get `d7_featureCounts.R` and `d7_featureCounts.sbatch` scripts
  - Edit both scripts and execute the `sbatch` script

Challenge Question

- What feature would you used to count reads for RNA-seq?

A. Gene

B. Exon

C. Transcripts

Challenge Question

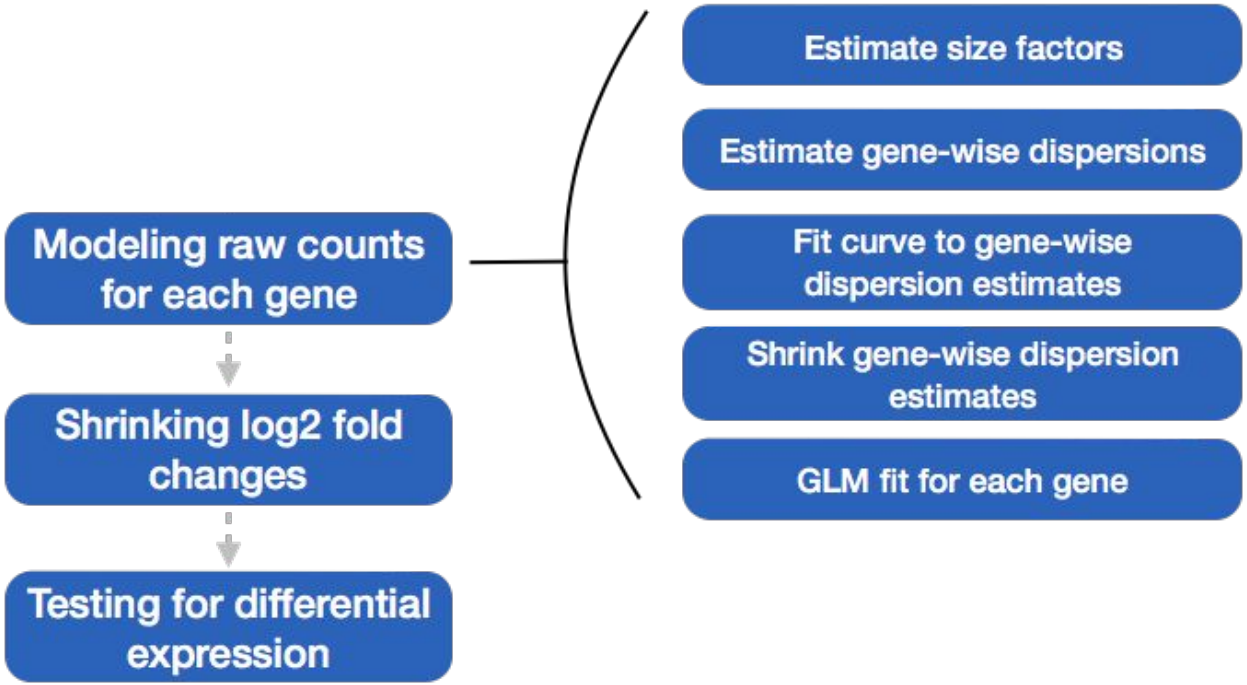
- What feature would you used to count reads for RNA-seq?

A. Gene

B. Exon ✓

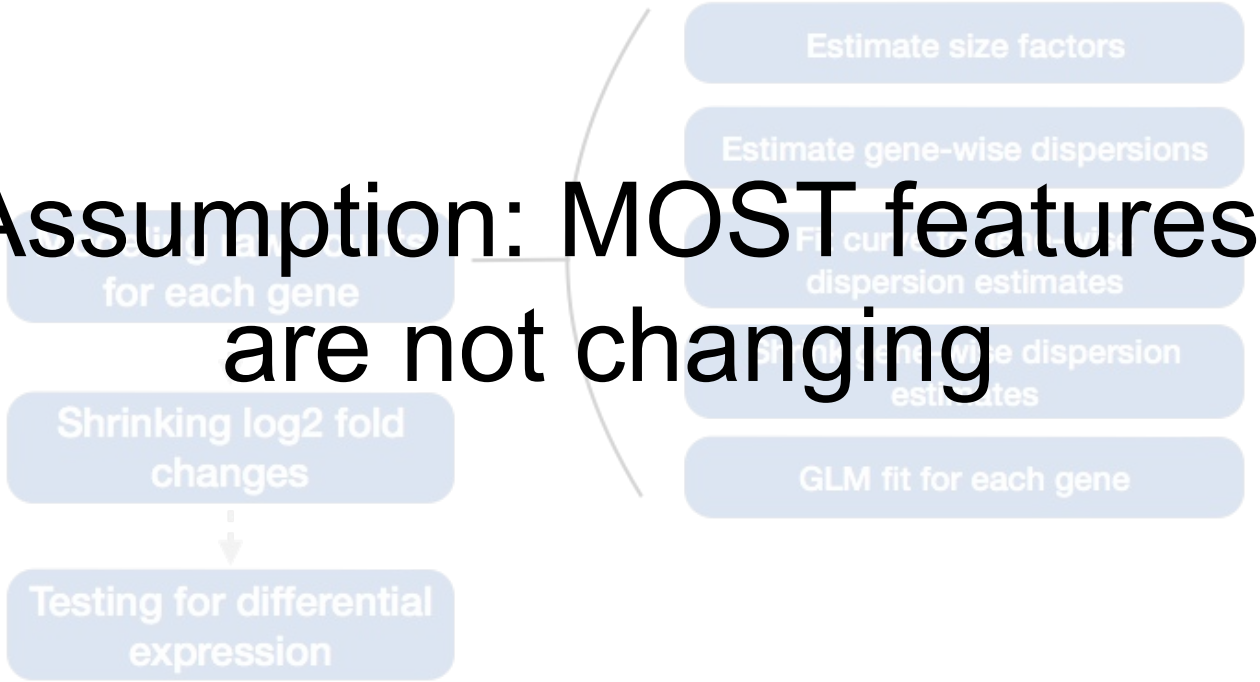
C. Transcripts

DESeq2 Recap



https://hbctraining.github.io/DGE_workshop_salmon/lessons/04_DGE_DESeq2_analysis.html

DESeq2 Recap

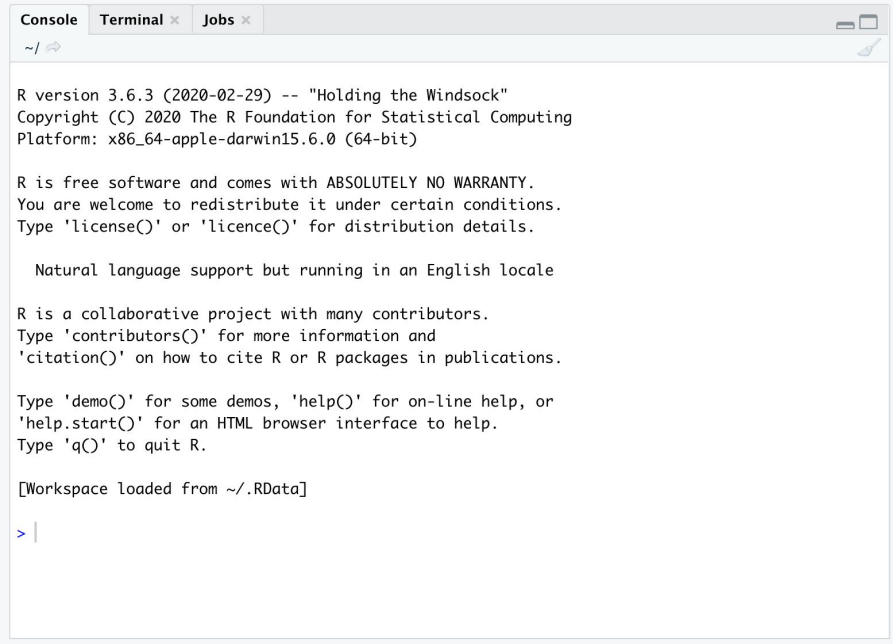


Assumption: MOST features are not changing

https://hbctraining.github.io/DGE_workshop_salmon/lessons/04_DGE_DESeq2_analysis.html

Run DESeq2…

- Follow DESeq2 worksheet
  - This will be run in an R console in RStudio



Challenge Question

- How would you run DESeq2 on the supercomputer?

Challenge Question

- How would you run DESeq2 on the supercomputer?
  - Install DESeq2 in your R packages directory
  - Make a conditions table that matches your count table
  - Run the R script through an sbatch script

Homework

- Run DESeq2 to explore differential expression with a different cell line